Automated Process for Accessing Vital Health Information at Census Tract Level

Hsiu-Hua Liao (1),* Paul Laymon (2), Kirk Shull (2)

(1) St. Louis County Department of Planning, Clayton, MO (2) Division of Biostatistics, South Carolina Department of Health and Environmental Control, Columbia, SC

Abstract

Recent emphasis on streamlined government and health care reform encourages community leaders to search for innovative ways to effectively manage their regions of responsibility. Gradually, geographic information system (GIS) technology is becoming a recognizable tool in the public health community for uses from intervention strategies to health care reform. One advantage of implementing GIS is that it can geographically locate personal health data through a geocoding process and allow examination of their spatial patterns. Georeferencing personal health data will greatly enhance decisions made by public health officials; however, it complicates the burden of protecting personal rights to confidentiality. One solution to the dilemma is to aggregate personal identities to a group of data where no identity will be revealed. This paper describes how this process was used to geocode vital health records and aggregate them to the census tract level. Data aggregation was accomplished through the Vital Health and Census Data Integration System (VHCDIS), an ARC/INFO-based GIS automation system. The primary objectives for the process were to promote personal privacy, automate health data aggregation of georeferenced vital records data, and improve national access to spatial health information.

Keywords: geocoding, vital health, automation, census data

Introduction

For centuries, health researchers have been using spatial locations, boundaries, and regions to determine the quality, quantity, and migration of epidemics. Overlaying quantitative graphics upon a map enables the viewer to realize potential information in an extremely clear manner. For example, the famous 1854 London cholera study conducted by Dr. John Snow has been hailed as the geographic benchmark for using maps in epidemiological studies.

Currently, the South Carolina Department of Health and Environmental Control (SCDHEC), Division of Biostatistics, presents spatial health information on the county level. County level data provide a wealth of information. At this macro scale, however, it is difficult for local health officials to adequately identify, analyze, and monitor health problems at a micro scale or community level. Hence, in 1989, the Johnson Wood Foundation authorized a grant for the SCDHEC's Vital Record Geographic Referencing System (VRGRS) and the University of South Carolina's School of Public Health to generate a feasibility study of georeferencing vital records data for the purpose of assisting

* Hsiu-Hua Liao, St. Louis County, Dept. of Planning, 41 S. Central Ave., Clayton, MO 63105 USA; (p) 314-615-3899; (f) 314-615-3729; E-mail: Hsui-Hua_Liao@co.st-louis.mo.us

public health assessments, surveillance, and health hazard evaluations at the community level. The main objectives for VRGRS were:

1. To implement a program that encoded the geographic residential location for births and deaths, and apply a geographic information system (GIS) as part of the statewide vital records system.
2. To demonstrate the application of location data in association with the TIGER (topologically integrated geographic encoding and referencing) system of the federal census of 1990.
3. To design and document the process in a way that facilitated expansion that complemented a statewide GIS for economic development.

The VRGRS project outcome ultimately determined that the processes, scientific techniques, and data were suitable enough to implement an informal GIS program within the Division of Biostatistics. Hence, in 1994 staff and equipment were selected to carry on the objectives of VRGRS and to establish the means to systematically georeference vital health data collected and stored at the Office of Vital Records.

Georeferencing provides an opportunity to examine health data and how they will be distributed over spatial domain; however, this also raises the issue of confidentiality. When the geographic resolution of data is fine enough to identify fewer than four addresses, the data are no longer tools of research, but tools to potentially target and expose individuals (1). The protection from inadvertent disclosure of individuals, households, establishments, or primary sampling units, especially in public use databases, is a concern of government health agencies. Even though confidentiality policies may vary among agencies, they must reflect the laws and regulations imposed upon personal data collection and dissemination activities (2). To date, there is not a minimum national threshold standard defining public or professional access to spatially referenced public health data.

In an attempt to promote spatially referenced public health confidentiality standards, the South Carolina Division of Biostatistics GIS Lab focused on the development of a statewide health information system capable of satisfying the wide range needs of health researchers. To develop such a system, the Biostatistics GIS Lab needed a geocoding system capable of converting large volumes of data with acceptable match rates. After a series of tests that included quality, cost, and turn around times, Geographic Data Technology (GDT) from Lebanon, New Hampshire, was chosen to perform the geocoding process.

Once the vital records health data were converted into individual points, the issue of confidentiality was solved by aggregating the data to the 1990 census tracts. The census tracts were chosen for two reasons. First, census tracts contained a volume of socioeconomic data. Thus, the aggregate vital records attribute information could be combined with the existing socioeconomic census data (e.g., mother's age extracted from the vital records would be stratified into the same categorical breakout as the female populace of the tracts, allowing calculation of statistical rates). Second, geographic boundaries are updated once every decade.

Working with voluminous vital records files proved to be tedious and time consuming. To streamline the process of generating public health data from these records, the Vital Health and Census Data Integration System (VHCDIS) was developed. In designing the system five requirements were determined:

- It must be flexible enough to be continuously improved.
- It must be a time saver.
- It must establish a national precedence for collecting health data.
- It must standardize data output.
- It must accurately aggregate health data to predetermined political boundaries (in this case the census tracts).

In its completed form, the VHCDIS offers national and local programs the ability to join aggregate vital records health data with existing socioeconomic census data as a tool for their respective surveillance and intervention strategies. The remaining point data derived from the geocoding process, which are treated with all the confidentiality of a paper certificate, are stored on a magnetic device for future use in very high resolution studies.

Background

Vital Health Statistics

Vital statistics for the United States are obtained from the official records of live births, deaths, fetal deaths, marriages, divorces, and annulments. These datasets have long been used as statistical measuring devices to identify qualitative and quantitative public health issues. The official recording of these events is the individual responsibility of each state and independent registration areas (District of Columbia, New York City, and territories). The federal government, without expressed constitutional authority to enact national vital statistics legislation, relies upon the states to establish laws and regulations to provide compatible methods of registration and data collection (3).

As public health issues continue to become more and more complex, demand for better vital statistics information increases. For this reason, updating data collecting, recording, and processing techniques to keep aligned with the rapidly evolving need becomes an increasingly important part of the vital statistics program. Improvement began in the 1950s with increased attention placed on improving the quality of vital statistics data to make them more useful and accessible. Interest in vital statistics expanded when the state and federal health and welfare officials began to look for pertinent and reliable statistics on which to base their political decisions. The registration certificates assumed a new role of importance as they were used as a source of credible national vital statistics by all levels of government, institutions, and the general public. The content of the information collected for vital records was expanded and methods to improve its quality and usefulness were added. As health and social issues became more complex, supplemental data sources were developed to augment and enrich the information obtained from the registration system.

Throughout the years the process of producing national vital statistics has shifted several times from one organizational unit of the federal government to another. The National Center for Health Statistics and the National Association for Public Health Statistics and Information System (NAPHSIS) have become recognized for handling health statistics and the associated information systems. NAPHSIS was organized to study and promote all matters relating to the registration of vital statistics. The 1995 revision of the association bylaws states:

This Association will foster discussion and group action on issues involving public health statistics, public health information systems, and vital records registration. The Association will provide standards and principles for administering public health statistics, public health information systems, and vital records registration. The Association will represent the States and Territories of the United States regarding these issues, and will serve as an advisory group to the Association of State and Territorial Health Officials (3).

With the increasing complexity of public health issues, federal and local health programs need to improve the process of collecting, storing, analyzing, and displaying community level epidemic information. For this reason, future focus on quality spatial vital records data will continue to grow at an exponential rate. Likewise, vital records recording programs tasked to increase the accuracy of vital statistics will continue to explore the development of new technologies, rethinking the use for these valuable resources.

Public Health and GIS

GIS technology has gradually been recognized by public health researchers as a powerful tool for analyzing health data. It provides an opportunity to integrate at least six disciplines (epidemiology, environmental health, geography, cartography, computer sciences, and statistics) for the study of the distribution and possible causes of diseases in population, and the targeting of interventions to improve the health of the population (4). Applications of GIS in the health field vary from the simple automated mapping of epidemiological data (5), to the sophisticated analysis of satellite images to demonstrate vector/environment relationship (6,7,8,9).

The simplified paradigm for implementing GIS technology in public health can be viewed in three phases: data source identification, GIS support system, and health planning (Figure 1). In the data source identification phase, data sources applicable to your cause are selected and converted into digital geography or "coverages." The data sources used in the Division of Biostatistics are environmental health hazards, health services, and socioeconomic and health data. Environmental health hazard data can be defined as any data pertaining to an environmental situation that may have a negative impact on the surrounding population. Health services data are those that identify sources of health correction. And, for the scope of this paper, socioeconomic and health data can be defined as data collected for the purpose of monitoring, tracking, and identifying social and health trends.

Once these data sources are converted into digital coverages, they can be stored, manipulated, analyzed, and displayed in a GIS. This collection of standardized data becomes the foundation for the third phase of health GIS implementation, the health planning phase. In this phase, the GIS becomes the knowledge base for analyzing health outcomes and supporting public health surveillance, where a diverse group of scientific disciplines converge to direct and discover local level health objectives.

The Centers for Disease Control and Prevention define public health surveillance as [t]he ongoing, systematic collection, analysis, and interpretation of health data essential to the planning, implementation, and evaluation of public health practice, closely integrated with the timely dissemination of these data to those who need to know. The final link of the surveillance chain is the
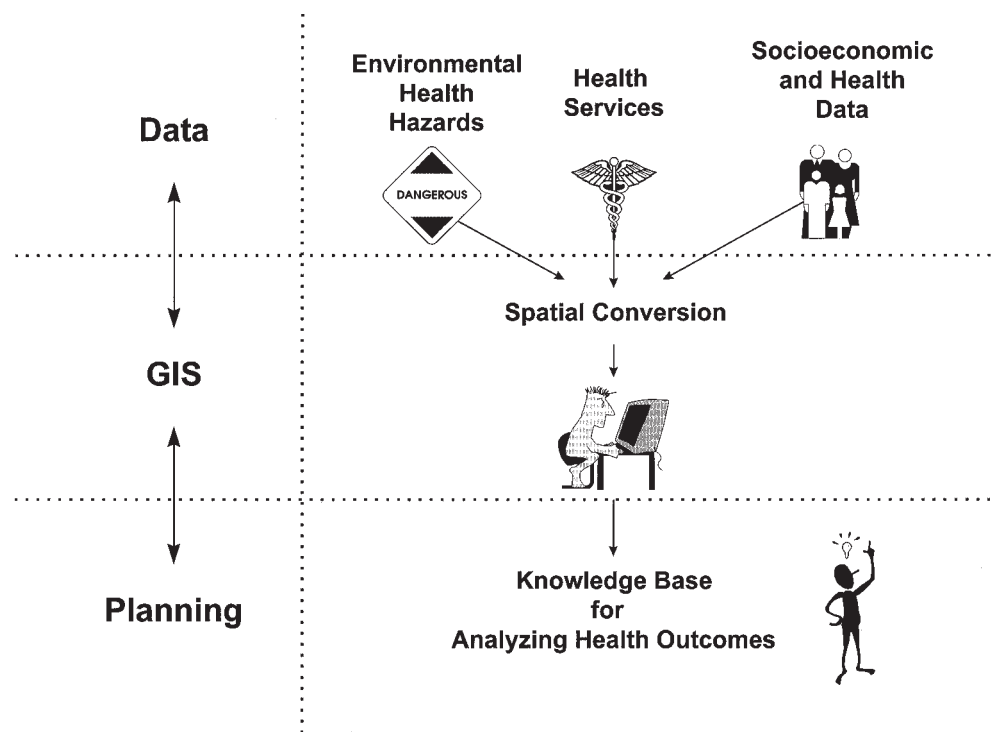
Figure 1 Paradigm for implementing GIS in public health.

application of these data to prevention and control. A surveillance system includes a functional capacity for data collection, analysis, and dissemination linked to public health programs (10).

Public health surveillance evolves with changes in science and technology. With the advent of computers, health surveillance has transformed from a primarily historical function to one that promotes timely analysis of data with appropriate responses to given health outcomes. Historically, the quality and quantity of data over a spatial domain were illusive and difficult to interpret. Today, using GIS technology as a tool we can streamline the processes needed to promote health and protect our environment.

Vital Health Data

In the state of South Carolina, vital health data are collected through official documents filed with the Office of Vital Records and the Public Health Statistics and Information System within the SCDHEC. Each year, the Division of Biostatistics publishes reports on vital statistics data for South Carolina live births, deaths, fetal deaths, marriages, divorces, and annulments that occurred during the previous year. These vital statistics are also available in publicly accessible format for public use. In the case of special requests for data, files and reports are generated and distributed by the Division of Biostatistics to those users who desire analysis different from those that are normally published.

VRGRS justified the use of GIS technology to improve the Division of Biostatistics'

capability of analyzing vital records data at an increased spatial resolution. Ultimately, this new technology functions around the process to geocode temporal vital records residence data of births and deaths in South Carolina. All births by residents were included, regardless of the state of occurrence, while South Carolina occurrences to non-residents were excluded. To support thematic mapping and GIS analysis, an attribute file identifying critical information about the birth and death events was generated and linked to the point by means of a common identifier.

Birth Data

In 1991, South Carolina began using a microcomputer software, Electronic Birth Pages (EBP), to improve the process of generating birth certificates and collecting newborn data for laboratory screening. The main function of the EBP system involves data entry and production of birth certificates. The end product is referred to as an EBC (electronic birth certificate).

To generate spatial information from vital records data, the residential address file is extracted from the mainframe dataset using the statistical software SAS (SAS Institute, Cary, NC). Variables used in geocoding include identification number, residential street address, city, state, zip code, and 4-digit zip code extension. These data undergo quality control measures to identify completeness and accuracy of the address information. To complete the dataset, an attribute record is captured as well. For example, the attribute file used for births includes identification number, county federal information processing standards (FIPS) code, age of mother, attendant at birth, birth weight, education level of the mother, month prenatal care began, number of prenatal care visits, race of the child, race of the mother, sex of the child, and year of birth. These chosen attributes were based on requests made by health districts and the Division of Epidemiology within SCDHEC. The attribute file was imported into ARC/INFO for the data aggregation process. Each variable was aggregated to the census tract level by race group (total, white, black, others, and unknown). Table 1 shows the classification of each variable.

Death Data

Death data were collected through death certificates filed by funeral homes. The funeral director, or person acting as such, is responsible for the completion of the death certificates, including all of the personal information from the family, and the medical portion of the certificate. This certificate is then sent to the county health department where it is screened for completeness. If the certificate is acceptable at the county level, the health department will forward the certificate to the SCDHEC's Office of Vital Records. The certificate is again checked for completeness, then personal data are coded and stored in the database.

For the residential address file, variables used in the geocoding process were synonymous with the birth data. For the attribute file, variables were temporally selected and causes of death were grouped into disease and non-disease type (Table 2). Causes of death are classified for purposes of statistical tabulation according to the *Manual of the International Statistical Classification of Diseases, Injuries, and Causes of Death* (11). In this process, only the underlying cause of death was selected for data aggregation.

Table 1   Classification of Birth Data: Live Birth, Low Weight Live Birth, and Very Low Weight Live Birth

A. LIVE BIRTH

RACE
  Total
  White
  Black
  Others
  Unknown

ATTENDANT
  Physician
  Certificated nurse midwife
  Other than physician, midwife,
    or self
  Self-attended
  Lay midwife, lay midhusband,
    registered lay midwife
  Nurse-midwife, graduate
    nurse-midwife
  Nurse, RN, OB nurse practi-
    tioner, physician's assistant
  D.O.
  Unreported or unknown

BIRTH WEIGHT
  <500 g
  500–999 g
  1,000–1,499 g
  1,500–1,999 g
  2,000–2,499 g
  2,500–2,999 g
  3,000–3,499 g
  3,500–3,999 g
  4,000–4,499 g
  4,500–4,999 g
  5,000 g
  Unreported

CHILD'S SEX
  Male
  Female
  Unreported

MOTHER'S AGE
  $\leq$13
  13
  14
  15
  16
  17
  18
  19
  20
  21
  22–24
  25–29
  30–34

35–39
40–44
$\geq$45
Unreported or unknown

MOTHER'S EDUCATION
  Elementary school
  1st grade
  2nd grade
  3rd grade
  4th grade
  5th grade
  6th grade
  7th grade
  8th grade
  9th grade
  10th grade
  11th grade
  12th grade
  1st year college
  2nd year college
  3rd year college
  4th year college
  Graduate school
  Technical school
  Unreported

MONTH PRENATAL CARE BEGAN
  No prenatal care
  Began at 1st month
  Began at 2nd month
  Began at 3rd month
  Began at 4th month
  Began at 5th month
  Began at 6th month
  Began at 7th month
  Began at 8th month
  Began at 9th month
  Began at 10th month
  Unreported

NUMBER OF PRENATAL CARE VISITS
  No prenatal care visit
  1–4 visits
  5 visits
  6–10 visits
  11–15 visits
  16 visits
  Unreported

**B. LOW WEIGHT LIVE BIRTH
(Birth Weight <2,500 g)**

RACE
  Total

White
Black
Others
Unknown

MOTHER'S AGE
  $\leq$13
  13
  14
  15
  16
  17
  18
  19
  20
  21
  22–24
  25–29
  30–34
  35–39
  40–44
  $\geq$45
  Unreported or unknown

**C. VERY LOW WEIGHT LIVE
BIRTH (Birth Weight <1,500 g)**

RACE
  Total
  White
  Black
  Others
  Unknown

MOTHER'S AGE
  $\leq$13
  13
  14
  15
  16
  17
  18
  19
  20
  21
  22–24
  25–29
  30–34
  35–39
  40–44
  $\geq$45
  Unreported or unknown

Table 2   Attributes of Death Data

DEATH	Benign and Unspecified Neoplasms	Others
RACE	**Cancer**	**Hernia and Intestinal Obstruction**
Total	Bladder	**Hypertension**
White	Brain and other nervous system	**Infectious and Parasitic**
Black	Female breast cancer	HIV/AIDS
Others	Male breast cancer	**Meningitis**
Unknown	Cervix uteri (female only)	**Nephritis, Nephrotic Syndrome, and Nephrosis**
AGE by race	Colon and rectum	**Pneumonia**
0	Corpus uteri (female only)	**All Other Diseases**
1–9	Esophagus	
10–19	Hodgkin's disease	**NON-DISEASE**
20–29	Kidney and renal pelvis	
30–39	Larynx	**Unintentional Injuries**
40–49	Leukemia	Drowning
50–59	Liver and intrahepatic bile duct	Falls
60–69	Lung and bronchus	Fire and flames
70–79	Melanoma of the skin	Firearms
80	Multiple myeloma	Motor vehicle accidents
AUTOPSY by race	Non-Hodgkin lymphoma	Poisoning by drugs and medicaments
Yes	Oral cavity and pharynx	Railway
No	Ovary	Others
BURIAL DISPOSITION by race	Pancreas	**Homicide and Legal Intervention**
Burial	Prostate	Assault by firearms
Cremation	Sarcoma	Assault by cutting and piercing
Donation	Stomach	Others
Hospital disposition	Testis (male only)	**Suicide**
Removal	Thyroid	Firearms
Other	**Cerebrovascular Disease**	Hanging, strangulation, and suffocation
Unreported	**Certain Cond. Originating in Perinatal Period**	Poisoning
SEX by race		Others
Female	**Chronic Liver Disease and Cirrhosis**	**Other External Causes**
Male	**Chronic Obstructive Pulmonary and Allied Cond.**	
CAUSE OF DEATH by race		
Disease	**Congenital Anomalies**	
Non-disease	**Diabetes**	
	Disease of Heart	
DISEASE	Ischemic	
Arteriosclerosis		

Geocoding Process

Geocoding is the process of linking a common location identifier such as address, site location, or building to a spatial and geographic database, such as one with census TIGER/Line files, that contains the locations of streets, the ranges of addresses found on each street segment, and the boundaries of political and administrative areas. Because the geographic database contains address ranges defining the beginning and ending address numbers that were assigned to a given street segment, coordinates (i.e., latitude and longitude) for any specific address location can be found through a linear

Table 3 GDT Geocoding Summary Information

Variable Definition		Notes
GDTPLUS4	4-digit zip code match	
GDTSAD	Street address match	
GDTCITY	City match	
GDTSTATE	State match	
GDTZIP	5-digit zip code match	
GDTSFIPS	State FIPS code match	
GDTCFIPS	County FIPS code match	
GDTTR90	Census tract match	
GDTBG90	Census block group match	
GDTXIN	Centroid type code	0 = Not a centroid (street address level match)
		1 = Zip + 4 centroid
		2 = Zip + 2 centroid
		X = 5 digit zip code centroid
		Blank = No centroid available
GDTSTAT	Match status code	B1 = Batch street address number match
		B2, B3 = Batch street intersection number match
		B5 = Batch matched to a street address on an alternate street name
		B6 = Batch matched to a placeholder
		B7 = Batch matched to a placeholder on an alternate street name
		10 = Not a valid 2-digit alpha state code
		11 = City in not found in the state
		12 = Incomplete or poorly formatted address
		14 = Could not find street name in the city
		15 = Could not match number, directional, or street type
		16 = Multiple addresses found in that range
		17 = Street intersection failure
		18 = City is present in the state but no named or addressed street are present

interpolation of the address number between the starting and ending address numbers assigned to the segment. Once the correct location is assigned, a location identifier is given a map coordinate and becomes a permanent geocode.

At SCDHEC, the statewide geocoding service is conducted by Geographic Data Technology (GDT). Data were address matched to the Dynamap/2000 series, which is a base map used and generated by GDT for the purpose of address matching (12). Table 3 shows the summary information on the geocoding process.

Address Standardization

To improve address accuracy and increase the geocoding match rate, StreetRite

software (Group 1 Software, Lanham, MD), is used to check and correct resident addresses. StreetRite compares the residual addresses against a database of every mailable address in the United States, deciphers inaccurate or incomplete addresses (e.g., misspelled street names and missing zip codes, cities, and states) and replaces them with the correct data. The addresses StreetRite is unable to match are then manually checked, and if an accurate match is found, the address is corrected.

Error Sources of Geocoding

Geocoding is a process of matching an address to a geographic location. The quality of the geocoding process is referred to as the geocoding match rate. An accurate geocoding match process depends on the quality of the address data and the geographic data. There are some errors inherent to the process, and in many cases it is difficult to determine how accurate the results are. It is important to document the potential error sources and understand how they could affect the quality and results of geocoding. The following are factors identified during our geocoding process that could affect the match rate.

Accuracy of Address

In geocoding vital health records data, the initial error is introduced when the individual or family providing information to the medical official are not made aware of the difference between mailing and residence addresses. Mailing addresses quite often are the post office box at the local post office, while the residence address is the street and street number at which the individual resides. New addresses created in the calendar year that do not exist in the current street/road database will also reduce the geocoding match rate.

Address Allocation

The geographic data used in the geocoding process contain a wealth of information about street locations, address ranges, and related information, but they are by no means complete. In urban areas, the percentage of street segments that contain address ranges may be as high as 90% or above. Some rural areas, however, do not contain any address ranges. Therefore, the geocoding matching rate will depend upon the study area.

Assigning Geographic Location

Geocoding is a comparison of each address in an event table with the address ranges in a target address database. When an event address matches the address range of a street segment, an interpolation is performed to locate and assign real-world coordinates to the event. For example, given a line with end point values of 0 and 100 and a street address of 50, the location of the address is estimated at the line's midpoint. The actual street address, however, may not be located at the midpoint of the line segment. During the aggregation process there is the potential for a small percentage of geocoded data to be captured in the wrong polygonal boundary. For instance, in Figure 2, Tract 1 will be assigned an erroneous value.
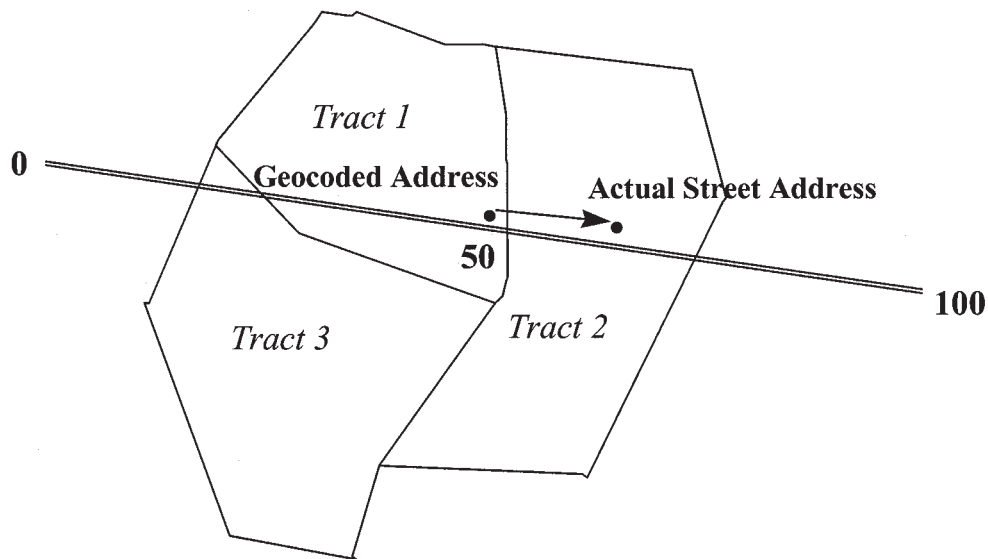
Figure 2   Illustration of potential error from assigning geographic location.

Automation System and Process

System Design

The primary goal of developing the VHCDIS is to aggregate vital health data to census tract level and generate publicly accessible database files. The system is designed to interact with users by generating aggregated information from different public health data. Figure 3 illustrates the general architecture of the VHCDIS. At this point, the system handles only birth and death records. In the future, as the need for geocoding health data increases, more components will be added to the system to handle different health information (such as cancer registry data).

System Resources and Implementation

The VHCDIS was developed using a GIS software, ARC/INFO (ESRI, Redlands, CA), on both Unix and NT platforms. A system supervisor in the form of an X-window graphical user interface (GUI) was used to provide user access to all the various components (birth data and death data) described previously. The GUI provides an interactive environment that facilitates user access to the components, as well as selection and execution of selected options. As described below, user navigation of the entire process is accomplished by appropriate selection from the window menu.

Table 4 summarizes the various steps involved in the automation process. The user first loads the system by opening and running the ARC/INFO software. At this point, the user is looking at the image shown in Figure 4. In this example, we will use the system to generate birth information. Therefore, the user can point and click on the Birth Certificate icon. As shown in Figure 5, twelve options are available for generating aggregated information on live births, low weight live births, and very low weight live births at the census tract level. The classification of each category is shown in Table 1.
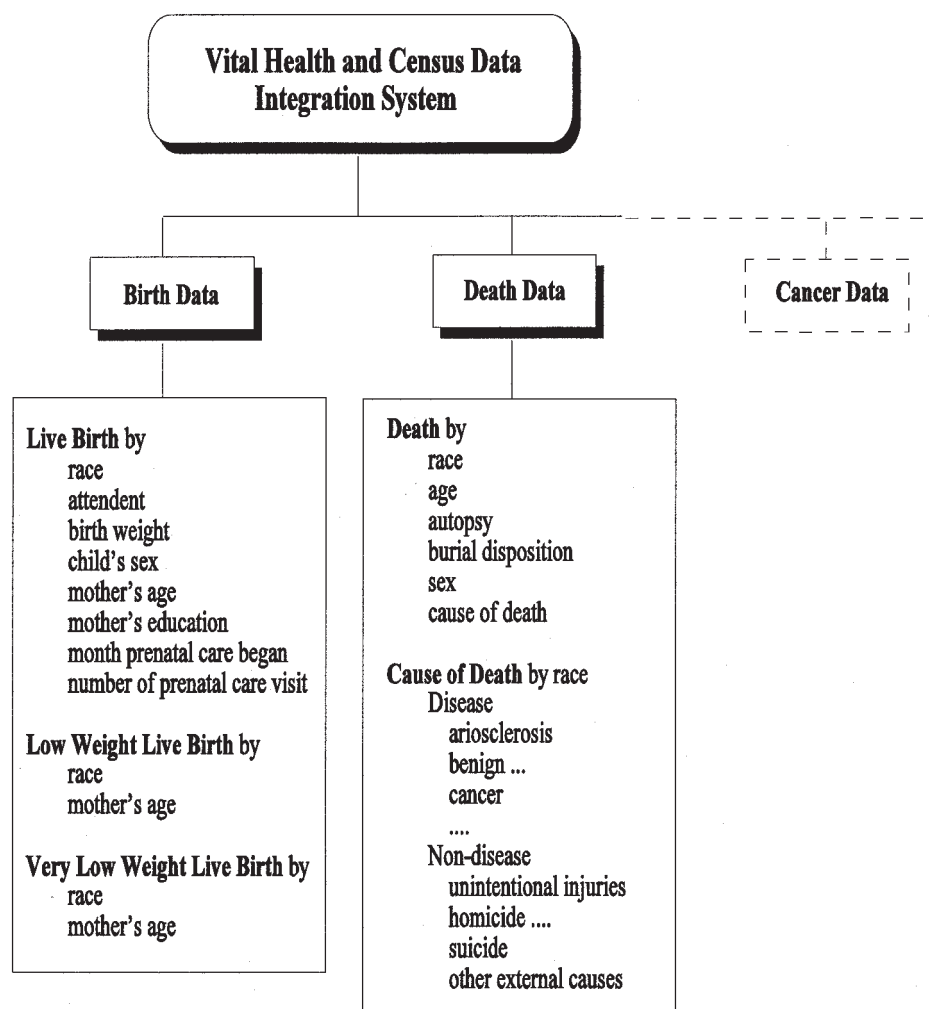
Figure 3   Architecture of the Vital Health and Census Data Integration System.

The user will make selections according to the information needed. For example, to generate live birth information by using the mother's age and race, the user will click the corresponding Select button to continue (Figure 6). In this menu, the user needs to supply two data files: birth data file (with all of the information shown in Table 1 plus the census county-tract number), and census tract data file (with census county-tract number only). When selecting a data file, a pop-up window will appear to help the user make the selection from existing data files.

After specifying the data file, the user can select field item names for each parameter (mother's age and mother's race from birth data; county-tract number from census data). The user should provide an output file name (e.g., lb95race.dbf) and define an item name for each classification shown in this menu. Because it is cumbersome to define the item names one by one, the user can select the USE DEFAULT button to use a

Table 4   Steps in Implementation of the Vital Health and Census Data Integration System

---

Component 1: Birth Data Aggregation

Step 1. Select **Birth Certificate** icon

Step 2. Select following variables accordingly:

Live births by mother's race

Live births by mother's age

Live births by mother's education

Live births by child's sex

Live births by prenatal care visits

Live births by prenatal care began

Live births by attendant at birth

Live births by birth weight

Low weight live births by mother's race

Low weight live births by mother's age

Very low weight live births by mother's race

Very low weight live births by mother's age

Example: select **live births by mother's age**

Step 3. Select birth attribute INFO file: **test95.dat**

Step 4. Select item name for mother's age: **MAGE**

Step 5. Select item name for mother's race: **MRACE**

Step 6. Select census tract INFO file: **test_cns.dat**

Step 7. Select item name for County-Tract number: **CNTR**

Step 8. Define output DBF file name: **lb95mage.dbf**

Step 9. Define individual item name or use default setting by selecting **USE DEFAULT**

Step 10. Select **DONE** to continue the process

Step 11. Select **CANCEL** to return to previous menu

Step 12. Select another variable

Component 2: Death Data Aggregation

Step 1. Select **Death Certificate** icon

Step 2. Select following variables accordingly:

Death by race

Death by age

Death by autopsy

Death by burial disposition

Death by sex

Death by cause of death

Example: select **death by age**

Step 3. Select death attribute INFO file: **dth95.dat**

Step 4. Select item name for age: **AGE**

Step 5. Select item name for race: **RACE**

Step 6. Select census tract INFO file: **test_cns.dat**

Step 7. Select item name for County-Tract number: **CNTR**

Step 8. Define output DBF file name: dth95age.dbf

Step 9. Define individual item name or use default by selecting **USE DEFAULT**

Step 10. Select **DONE** to continue the process

Step 11. Select **CANCEL** to return to previous menu

Step 12. Select another variable

---

default name for each item. After this, the user can click DONE to continue the generation process. When this process is finished, the user clicks CANCEL to return to the previous menu (Figure 5) to select other parameters for data aggregation. Outputs from the aggregation process are dBASE files (.dbf) that can be imported to other database software or ArcView (ESRI, Redlands, CA) to review.
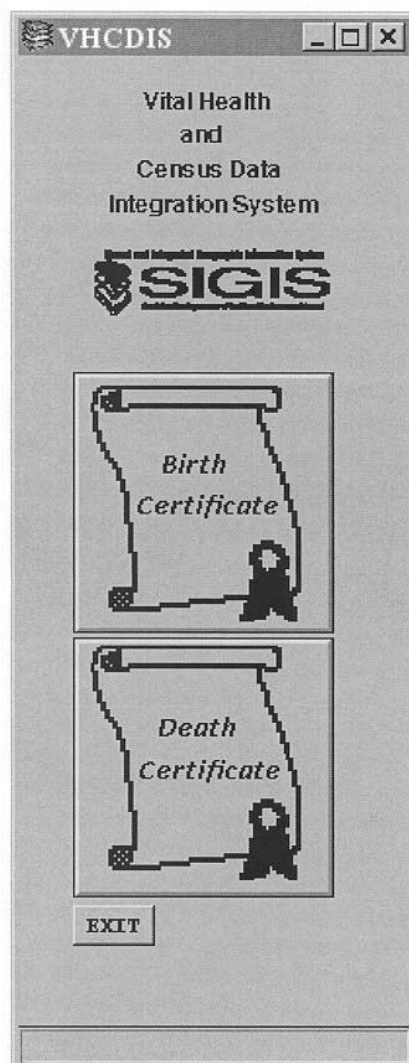
Figure 4   Main menu of the Vital Health and Census Data Integration System.

In this example, an output file (lb95race.dbf) was generated for live birth by mother's age and by race at the census tract level. To display the information, the user first runs the ArcView software, opens a graphic window, and adds the census tract coverage (or shape file) as a new theme. The user then needs to add the output file (lb95race.dbf) as a Table file and open the census tract attribute table. After opening the two tables, the user can perform the spatial join function to join both files and select different classified information to display (Figure 7).

Conclusion and Ongoing Process

GIS technology is emerging as a useful tool in public health studies. The technology

Figure 5   Birth information aggregation menu.

allows for storage and manipulation of large and multi-faceted datasets while maintaining the spatial integrity of each data collection or reporting location. As such, the technology gives rise to intensive investigation of the spatial relationship between variables and outcomes necessary to health risk assessment. Therefore, the key to successful application of GIS technology in the public health field is to understand what GIS functions should be used, what the limitations are, and how we should apply it appropriately to benefit research and assist officials with intervention strategies and health prevention.

In South Carolina, vital health data are collected each year. The need to access spatial health information is increasing as public health officials and researchers see the importance of analyzing spatial patterns of vital health data. Hence, creating a system to assist in standardizing data transformation from individual geocoded confidential health data to non-confidential data is needed.

This paper described an interactive data integration system, the Vital Health and Census Data Integration System (VHCDIS), developed and designed through the use of GIS technology for transforming geocoded confidential health data (birth and death) to non-confidential census health information. Vital health data were linked to census data through the geocoding process. By aggregating geocoded vital health data to

Figure 6   Generating information of live births by mother's age and by race.

census tracts, the output from VHCDIS will be publicly accessible and can be analyzed concurrently with other existing census socioeconomic data.

This report describes a project with the primary goal of developing a system to assist ongoing and systematic collection of health data and disseminate these data to public health officials and researchers for planning, implementing, and evaluating public health practice. Currently, SCDHEC is using the system to develop census health information from South Carolina birth and death data and will continue the process in the future. For the VHCDIS, many improvements and extensions are still underway. For
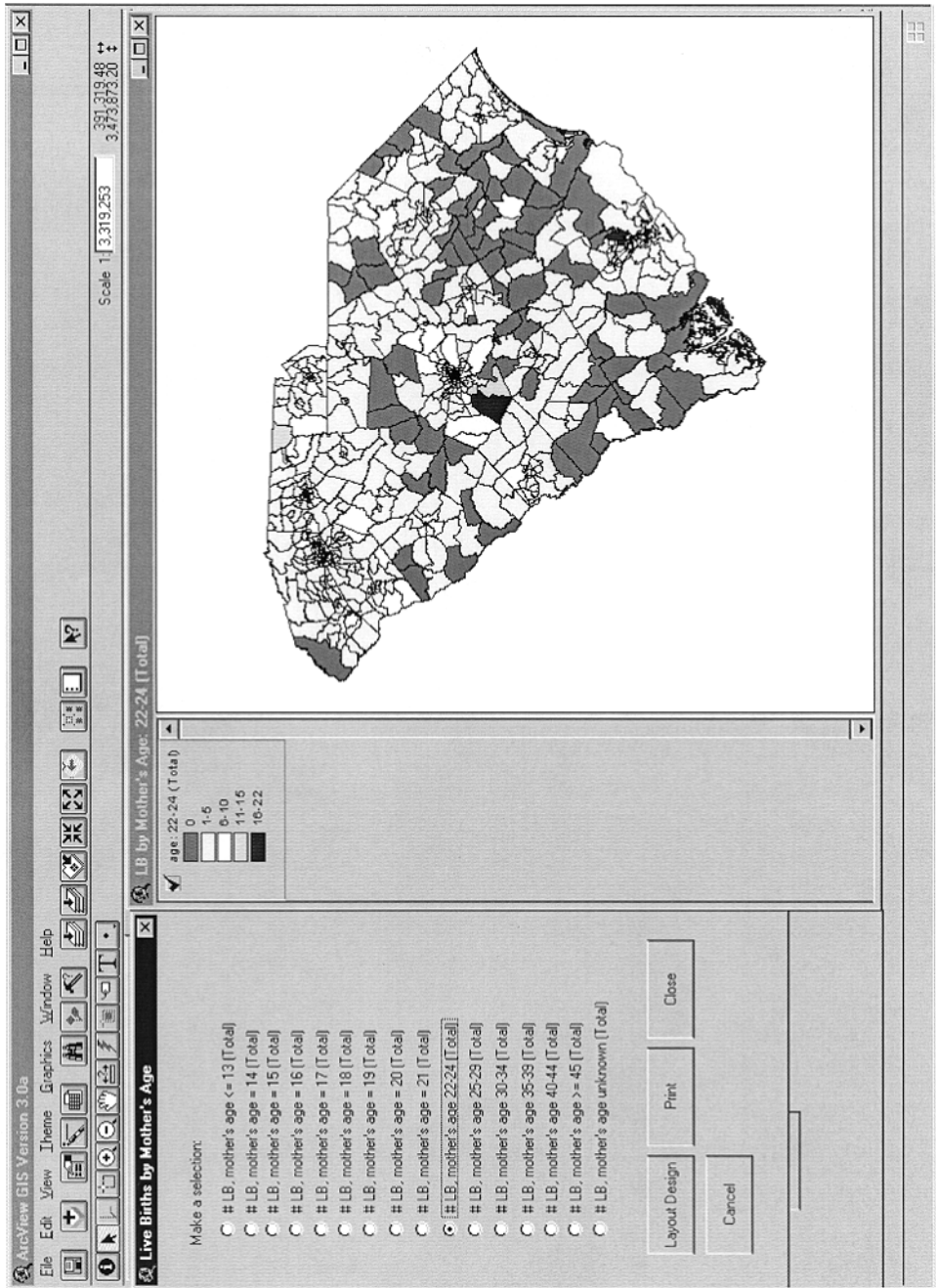
Figure 7   Output data display in ArcView for live births by mother's age and by race.

example, the system is being extended to accommodate infant death data and include cancer registry data next year. Additionally, the system can be extended to census block group levels and possibly to census block levels. In general, the VHCDIS in its present form is a sufficiently realistic demonstration of the flexibility of GIS technology and its ability to aggregate and process large volumes of health data.

References

1. Alpert S, Haynes KE. 1994. Privacy and the intersection of geographical information and intelligent transportation systems. In: *Proceedings of the Conference on Law and Information Policy for Spatial Database.* Tempe, AZ: National Center for Geographic Information and Analysis/Center for the Study of Law, Science, and Technology, Arizona State University College of Law. 198–211.

2. Croner MC, Sperling J, Broome FR. 1996. Geographic information system (GIS): New perspectives in understanding human health and environmental relationships. *Statistics in Medicine* 15:1961–77.

3. National Center for Health Statistics. 1995. *US vital statistics system: Major activities and developments, 1950–95.* Hyattsville, MD: US Department of Health and Human Services, Centers for Disease Control and Prevention. DHHS Publication No. (PHS) 97-1003.

4. Feinleib M. 1997. The use of computer mapping in monitoring the nation's health. In: *Proceedings from the International Symposium on Computer Mapping in Epidemiology and Environmental Health.* Geneva: World Health Organization. 1–3.

5. Pyle GF. 1994. Mapping tuberculosis in the Carolinas. *Sistema Terra* 3(1):22–3.

6. Hugh-Jones M. 1997. Applications of remote sensing to the identification of the habitats of parasites and disease vectors. *Parasitology Today* 5(8):244–51.

7. Malone JB, Fegker DP, Loyacano AF, Zukowski SH. 1992. Use of LANDSAT MSS imagery and soil type in geographic information system to assess site-specific risk of fascioliasis on red river basin farms in Louisiana. Reprinted from *Tropical Veterinary Medicine: Current Issues and Perspectives*. In: *Annals of the New York Academy of Sciences* 635:389–97.

8. Perry BD, Kruska R, Lessard P, Norval RAI, Kundert K. 1991. Geographic information systems for the development of tick-bone disease control strategies in Africa. *Preventive Medicine* 11:261–68.

9. Rogers DJ, Rabdikog SE. 1991. Mortality rates and population density of tsetse flies correlated with satellite imagery. *Nature* 351:739–41.

10. Centers for Disease Control and Prevention (CDC). 1988. *CDC surveillance update.* Atlanta: CDC.

11. World Health Organization (WHO). 1977–78. *Manual of the international statistical classification of diseases, injuries, and causes of death: based on the recommendations of the Ninth Revision Conference, 1975, and adopted by the Twenty-Ninth World Health Assembly.* Geneva: WHO.

12. Geographic Data Technology (GDT). 1997. *Dynamap/2000 7.2 user manual.* Lebanon, NH: GDT.